

第 1 章 数据库系统概述

数据库技术从 20 世纪 60 年代中期产生至今, 已经发展成为计算机科学的重要分支, 还带动了一个巨大的软件产业。数据库技术是信息系统的一个核心技术, 是使用计算机对各种信息、数据进行收集、管理的必备知识, 是计算机类学科, 包括计算机科学与技术、软件工程、网络工程等专业的重要课程, 也是许多其他非计算机类专业的必修或选修课程。数据库技术所研究的问题就是如何科学地组织和存储数据, 如何高效地获取和处理数据。

本章主要介绍数据库的基本知识, 包括数据库的发展历史、概念描述、体系结构等。

1.1 数据库系统简介

数据库技术是随着信息社会对数据处理任务的需要而产生的。随着社会对数据处理任务的要求不断提高, 数据库也随之产生并不断发展。数据库的诞生和发展给计算机信息管理带来了一场巨大的革命。数据库技术从问世到现在, 形成了坚实的理论基础、成熟的商业产品和广泛的应用领域, 吸引了越来越多的研究者加入。对于一个企业、一个国家来说, 数据库的建设规模、数据库信息量的大小和使用频度已经成为衡量这个企业、这个国家信息化程度的重要标志。

1.1.1 数据库技术的发展历史

在 20 世纪 60 年代, 随着数据处理自动化的发展, 数据库技术应运而生。在计算机应用领域中, 数据处理越来越占主导地位, 数据库技术的应用也越来越广泛。数据库是数据管理的产物。数据管理是数据库的核心任务, 内容包括对数据的分类、组织、编码、储存、检索和维护。从数据管理的角度看, 数据库技术到目前共经历了人工管理阶段、文件系统阶段和数据库系统阶段。

1. 人工管理阶段

人工管理阶段是指计算机诞生的初期, 即 20 世纪 50 年代后期之前。这个时期的计算机主要用于科学计算。从硬件看, 起初没有磁盘等直接存取的存储设备, 后来出现了磁带; 从软件看, 没有操作系统和管理数据的软件, 数据处理方式是批处理。那时的数据管理非常简单, 通过大量的分类、比较和表格绘制的机器运行数百万穿孔卡片或读写磁带来进行数据的处理和存储。

这个时期的数据管理具有以下 4 个特点。

(1) 数据基本不保存

该时期的计算机主要应用于科学计算, 由于技术限制, 一般不需要将数据长期保存, 只是在计算某一课时将数据输入, 用完后不保存原始数据, 也不保存计算结果。

(2) 没有对数据进行管理的软件系统

程序员不仅要规定数据的逻辑结构, 还要在程序中设计物理结构, 包括存储结构、存取方法、输入/输出方式等。因此, 程序中存取数据的子程序随着存储的改变而改变, 数据与程序不具有 consistency。卡片或磁带都只能顺序读取。

(3) 没有文件的概念

数据的组织方式必须由程序员自行设计。数据是面向应用的，一组数据只能对应一个程序。即使两个程序用到相同的数据，也必须各自定义、各自组织，数据无法共享、无法相互利用和相互参照，从而导致程序与程序之间有大量重复的数据。因此，程序与程序之间有大量的冗余数据，数据不能共享。

(4) 数据不具有独立性

当数据的逻辑结构或物理结构发生变化后，必须对应用程序做相应的修改，这就加重了程序员的负担。

人工管理阶段应用程序与数据之间的对应关系，如图 1-1 所示。

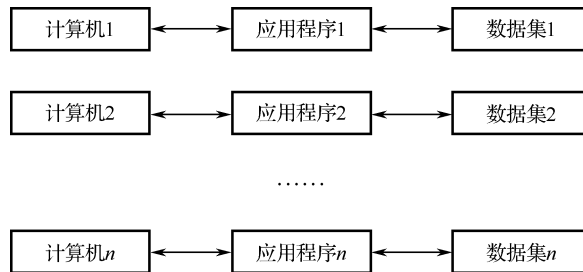


图 1-1 人工管理阶段应用程序与数据之间的对应关系

2. 文件系统阶段

20 世纪 50 年代后期至 60 年代中期，这时在硬件方面，外存储器有了磁盘、磁鼓等直接存储的存储设备；在软件方面，操作系统中已经有了专门用于管理数据的软件，称为文件系统。处理方式上不仅有了批处理，还有能够联机的实时处理。

这个时期的数据管理具有以下 4 个特点。

(1) 数据可以长期保存

由于计算机大量用于数据处理，经常对文件进行查询、修改、插入和删除等操作，因此数据需要长期保留在外存储器中，便于反复操作。

(2) 由文件系统管理数据

操作系统提供了文件管理功能和访问文件的存取方法，程序与数据之间有了数据存取的接口，程序可以通过文件名和数据打交道，不必再寻找数据的物理存放位置。至此，数据有了物理结构和逻辑结构的区别，但此时程序与数据之间的独立性尚不充分。

(3) 文件的形式已经多样化

由于已经有了直接存取的存储设备，文件也就不再局限于顺序文件，还有了索引文件、链表文件等，因而对文件的访问可以是顺序访问，也可以是直接访问。

(4) 数据具有一定的独立性

文件系统中的文件是为某一特定应用服务的，因此，系统不容易扩充，仍旧有大量的冗余数据，数据共享性差，独立性差。

文件系统阶段应用程序与数据之间的对应关系，如图 1-2 所示。

3. 数据库系统阶段

数据库系统的萌芽出现于 20 世纪 60 年代。当时的计算机开始广泛地应用于数据管理，对数据的共享提出了越来越高的要求。传统的文件系统已经不能满足人们的需要。能够统一管理

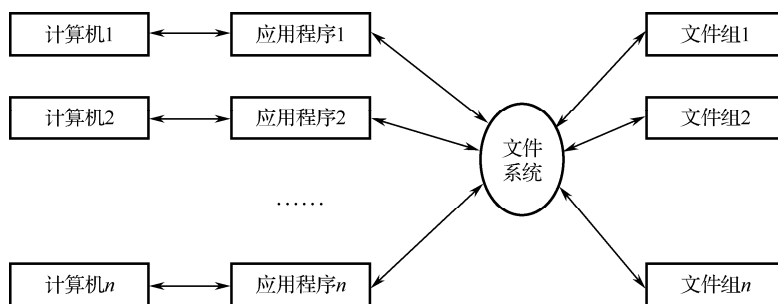


图 1-2 文件系统阶段应用程序与数据之间的对应关系

和共享数据的数据库管理系统便应运而生。

在这个阶段中，数据库中的数据不再是面向某个应用或某个程序，而是面向整个企业或整个应用的，处理的数据量急剧增长。这时在硬件方面，磁盘容量越来越大，读写速度越来越快；在软件方面，编制越来越复杂，功能越来越强大。处理方式上，联机处理要求更多。

这个时期的数据管理具有以下 4 个特点。

(1) 采用复杂的结构化的数据模型

数据库系统不仅要描述数据本身，还要描述数据之间的联系。这种联系是通过存取路径来实现的。数据结构化是数据库与文件系统的根本区别。

(2) 较高的数据独立性

数据与程序彼此独立，数据存储结构的变化尽量不影响用户程序的使用。数据与程序的独立把数据的定义从程序中分离出去，加上数据由数据库管理系统管理，从而简化了应用程序的编制，减轻了程序员的负担。

(3) 最低的冗余度

数据库系统中的重复数据被减少到最低程度，这样，在有限的存储空间内可以存放更多的数据，并减少存取时间。数据冗余度低，共享性高，易于扩充。

(4) 数据控制功能

数据库系统具有数据的安全性，以防止数据的丢失和被非法使用；具有数据的完整性，以保护数据的正确、有效和相容；具有数据的并发控制，避免并发程序之间的相互干扰；具有数据的恢复功能，在数据库被破坏或数据不可靠时，系统有能力把数据库恢复到最近某个时刻的正确状态。

数据库系统阶段的程序与数据之间的关系如图 1-3 所示。

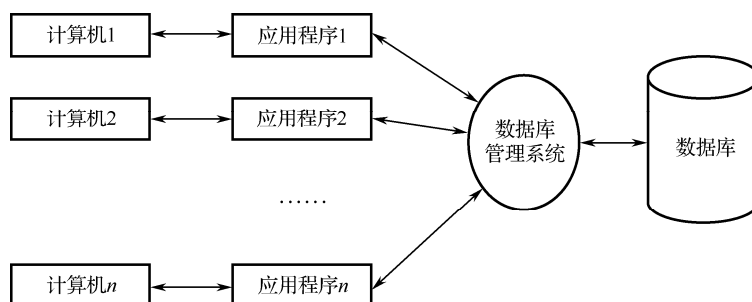


图 1-3 数据库系统阶段应用程序与数据之间的对应关系

事实上，只要有大量的信息需要处理，需要大量数据支持工作，就可以使用数据库技术。目前，数据库技术应用到了社会生活中几乎所有的领域，在金融业、航空业、学校、人力资源和军事等领域，数据库已经成为了不可或缺的组成部分。

随着科学技术的不断进步，各领域对数据库技术提出了更多的需求，现有数据库已经不能完全满足需求，于是新一代数据库孕育而生。新一代的数据库支持多种数据模型，并和诸多新技术相结合，广泛应用于更多领域。

总之，随着科学技术的发展，计算机技术不断应用到各行各业，数据存储不断膨胀，对未来的数据库技术将会有更高的要求。

1.1.2 数据库系统的基本概念

数据库系统作为信息系统的核心和基础，涉及一些常用的术语和基本概念。

1. 数据

数据（Data）是数据库中存储的基本对象。数据在大多数人的头脑中的第一反应就是数字，例如 50，-1000，¥90 等。其实数字只是最简单的一种数据。广义数据的种类很多，不仅仅指的是具体的数字和文字，还包括图形、声音、人的身体状况记录、计算机运行情况等，这些形式的数据经过数字化后都可以存储到计算机中。因此，数据是人们用各种物理符号，把信息按一定格式记载下来的有意义的符号组合。

例如，学生王纯峰的计算机网络考试成绩为 95 分，95 就是数据。学生王纯峰还有学号、姓名、性别、出生日期、学院名称等信息。所以可这样来描述学生王纯峰：

（2017130215，王纯峰，男，1998-1-10，计算机学院）

这些描述也是数据，它们都可以经过数据化处理后被计算机识别。通过这些数据可以掌握该学生的基本信息，但无法正确理解数据的含义，例如，数据 2017130215 的含义是学号还是身份证号，从数据中无法知道。可见，数据的形式还不能完全表达其内容，需要解释。因此，数据和关于数据的解释是不可分的。数据的解释是指对数据含义的说明，称为数据的语义。数据与其语义是不可分的。

2. 数据库

数据库（DataBase，简称 DB）即存放数据的仓库，这个仓库就是计算机存储设备。数据库中的数据并不是简单的堆积，数据之间相互关联。严格地讲，数据库是长期存储在计算机内、有组织的、可共享的大量数据的集合。数据库中的数据按照一定的数据模型组织、描述和存储，具有较小的冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。

3. 数据库管理系统

数据库管理系统（DataBase Management System，简称 DBMS）是专门用于管理数据库的计算机系统软件，介于应用程序与操作系统之间，是一层数据管理软件。数据库管理系统能够为数据库提供数据的定义、建立、维护、查询和统计等操作功能，并完成对数据完整性、安全性进行控制的功能。

现今广泛使用的数据库管理系统有微软公司的 Microsoft SQL Server、Access，甲骨文公司的 ORACLE、MySQL，IBM 公司的 DB2、Informix 等。

数据库管理系统的主要功能包括：数据库定义功能、数据库存取功能、数据库管理功能、数据库建立维护功能。

4. 数据库系统

数据库系统（DataBase System，简称 DBS）是指在计算机系统中引入了数据库后的系统，

由计算机硬件、数据库、数据库管理系统、应用程序和用户构成，即由计算机硬件、软件和使用人员构成。数据库系统是一个计算机应用系统。

计算机硬件是数据库系统的物质基础，是存储数据库及运行数据库管理系统的硬件资源，主要包括主机、存储设备、I/O 通道等，以及计算机网络环境。

软件主要包括操作系统以及数据库管理系统本身。此外，为了开发应用程序，还需要各种高级语言及其编译系统，以及各种以数据库管理系统为核心的应用开发工具软件。

数据库管理系统是负责数据库存取、维护和管理的系统软件，是数据库系统的核心，其功能的强弱是衡量数据库系统性能优劣的主要指标。

数据库中的数据由数据库管理系统进行统一管理和控制，用户对数据库进行的各种操作都是由数据库管理系统实现的。

应用程序 (Application) 是在数据库管理系统的基础上，由用户根据应用的实际需要开发的、处理特定业务的应用程序。

用户 (User) 是指管理、开发、使用数据库系统的所有人员，通常包括数据库管理员 (DBA)、应用程序员和终端用户。

综上所述，在数据库系统中，数据库中包含的数据是存储在存储介质上的数据文件的集合；每个用户均可使用其中的部分数据，不同用户使用的数据可以重叠，同一组数据可以为多个用户共享；数据库管理系统为用户提供对数据的存储组织、操作管理功能；用户通过数据库管理系统和应用程序实现数据库系统的操作与应用。数据库强调的是数据，数据库管理系统强调的是系统软件，数据库系统强调的是系统。

数据库管理系统的地位如图 1-4 所示。

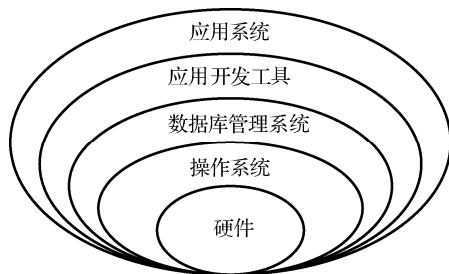


图 1-4 数据库管理系统的地位

1.2 数据库系统结构

数据库系统虽然是一个庞大、复杂的系统，但它有一个总的框架。虽然数据库系统软件产品众多，建立在不同的操作系统之上，但从数据库系统管理角度看，数据库系统通常采用三级模式结构，这是数据库管理系统内部的系统结构。

1.2.1 数据库系统的三级模式结构

数据库系统通常采用三级模式结构：外模式、模式和内模式，如图 1-5 所示。

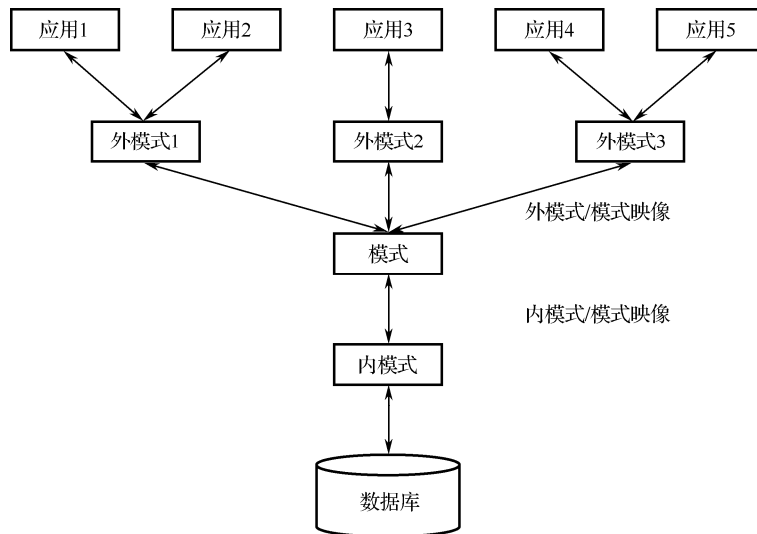


图 1-5 数据库系统的三级模式结构

1. 模式

模式也称为逻辑模式，是对数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图。

模式是数据库系统模式结构的中间层，既不涉及数据的物理存储细节和硬件环境，也与具体的应用程序和开发工具无关。模式实际上是数据库数据在逻辑级上的视图，一个数据库只能有一个模式。数据库模式以某一种数据模型为基础，综合考虑了所有用户的需求，并将这些需求有机地整合成一个逻辑整体。模式只是对数据库结构的一种描述，而不是数据库本身，是装配数据的一个框架。例如，数据记录由哪些数据项构成，数据项的名字、类型、取值范围等，而且要定义数据之间的联系，定义与数据有关的安全性、完整性要求。

数据库系统提供模式描述语言（模式 DDL）来严格地表示这些内容。

2. 外模式

外模式也称为子模式或用户模式，是数据库用户看到的数据视图，是与某一应用有关的数据的逻辑表示。

外模式是模式的子集，它是每个用户的数据视图。由于不同的用户其需求不同，看待数据的方式不同，对数据的要求不同，使用的程序设计语言也不同，因此不同用户的外模式描述是不同的。即使是模式中的同一数据，在外模式中的结构、类型、保密级别等方面都可以不同。一个数据库可以有多个外模式。

数据库系统提供外模式描述语言（外模式 DDL）来描述用户数据视图。

3. 内模式

内模式也称为存储模式，是数据在数据库系统内部的表示或底层描述，即对数据库物理结构和存储方式的描述。

一个数据库只能有一个内模式。例如，记录的存储方式是顺序存储、链式存储还是按哈希（Hash）方式存储；索引按照什么方式组织；数据是否压缩，是否加密等。

数据库系统提供内模式描述语言（内模式 DDL）来描述数据库的物理存储。

1.2.2 数据库系统的二级映像

数据库系统的三级模式是对数据的三个抽象级别，它把数据的具体组织留给 DBMS 管理，使用户能逻辑地、抽象地处理数据，而不必关心数据在计算机中的表示和存储。为了实现这三个层次上的联系和转换，数据库系统在这三级模式中提供了两层映像：外模式/模式的映像和模式/内模式的映像。

1. 外模式/模式的映像

模式描述的是数据的全局逻辑结构，外模式描述的是数据的局部逻辑结构。对于每一个外模式，数据库都有一个外模式/模式的映像，它定义并保证了外模式与数据模式之间的对应关系。这些映像定义通常包含在各自的外模式中。

当模式改变时（例如，增加新的关系、新的属性、改变属性数据类型等），外模式/模式的映像要进行相应的改变（由 DBA 负责），以保证外模式保持不变。应用程序是根据数据的外模式编写的，因此应用程序不必修改，保证了数据与程序的逻辑独立性，即数据的逻辑独立性。

2. 模式/内模式的映像

数据库的内模式依赖于它的全局逻辑结构，即模式。因为一个数据库只有一个模式，也只有一个内模式，所以模式/内模式的映像也是唯一的。它定义并保证了数据的逻辑模式与内模式之间的对应关系。

当数据库的存储结构改变了，模式/内模式的映像也必须进行相应的修改（仍由 DBA 负责），使得模式保持不变，保证了数据与程序的物理独立性，即数据的物理独立性。

正是由于上述二级映像功能，才使得数据库系统中的数据具有较高的逻辑独立性和物理独立性。二级映像保证了数据库外模式的稳定性，从而从底层保证了应用程序的稳定性。数据与程序之间的独立性使得数据的存取由 DBMS 管理，用户不必考虑存取路径等细节，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

1.3 习题

1. 数据库的发展历史分为哪几个阶段？各有什么特点？
2. 简述数据、数据库、数据库管理系统、数据库系统的概念。
3. 试述数据库系统的三级模式结构和二级映像的特点。
4. 什么是数据与程序的逻辑独立性？什么是数据与程序的物理独立性？
5. 举例说明，在实际工作生活中有哪些单位部门使用数据库？这些数据库所起的作用如何？